

CLAIMS

WHAT IS CLAIMED IS:

1. A method for processing data in a distributed architecture, the method comprising the steps of:
 - gathering information content from at least one repository according to a predetermined schedule;
 - registering the information content;
 - assigning the information content at least one document identifier;
 - transmitting at least one work request regarding at least a portion of the information content to a first work queue;
 - processing the at least one work request;
 - transmitting the at least a portion of the information content to a second work queue; and
 - processing the at least a portion of the information content.
2. The method of claim 1, further comprising the step of:
 - converting the at least a portion of the information content to a meta-document representation of the information content.
3. The method of claim 2, wherein the meta-document representation comprises extensible markup language (XML) format.
4. The method of claim 2, further comprising the step of:
 - analyzing the meta-document representation.
5. The method of claim 2, further comprising the step of:
 - indexing the meta-document representation.

6. The method of claim 1, further comprising the step of:
generating progress statistics regarding the step of processing the at least a portion of the information content.

7. The method of claim 6, further comprising the step of:
transmitting the progress statistics to a third work queue.

8. The method of claim 1, wherein the first work queue and the second work queue share access to a central data structure.

9. The method of claim 8, wherein access is shared via a CORBA service.

10. The method of claim 8, wherein the data structure represents at least one of a metrics history and taxonomy regarding the information content.

11. A system for processing data in a distributed architecture, the system comprising:
an information content gathering module that gathers information content from at least one repository according to a predetermined schedule;
a registering module that registers the information content;
an assigning module that assigns the information content at least one document identifier;
a work request transmitting module that transmits at least one work request regarding at least a portion of the information content to a first work queue;
a work request processing module that processes the at least one work request;
an information content transmitting module that transmits the at least a portion of the information content to a second work queue; and
an information content processing module that processes the at least a portion of the information content.

12. The system of claim 11, further comprising:

a converting module that converts the at least a portion of the information content to a meta-document representation of the information content.

13. The system of claim 12, wherein the meta-document representation comprises extensible markup language (XML) format.

14. The system of claim 12, further comprising:

an analyzing module that analyzes the meta-document representation.

15. The system of claim 12, further comprising:

an indexing module that indexes the meta-document representation.

16. The system of claim 11, further comprising:

a generating module that generates progress statistics regarding the processing of the at least a portion of the information content.

17. The system of claim 16, further comprising:

a progress statistics transmitting module that transmits the progress statistics to a third work queue.

18. The system of claim 11, wherein the first work queue and the second work queue share access to a central data structure.

19. The system of claim 18, wherein access is shared via a CORBA service.

20. The system of claim 18, wherein the data structure represents at least one of a metrics history and taxonomy regarding the information content.

21. A system for processing data in a distributed architecture, the system comprising:
gathering means for gathering information content from at least one repository according to a predetermined schedule;
registering means for registering the information content;

assigning means for assigning the information content at least one document identifier;
work request transmitting means for transmitting at least one work request regarding at
least a portion of the information content to a first work queue;
work request processing means for processing the at least one work request;
information content transmitting means for transmitting the at least a portion of the
information content to a second work queue; and
information content processing means for processing the at least a portion of the
information content.

22. The system of claim 21, further comprising:

converting means for converting the at least a portion of the information content to a
meta-document representation of the information content.

23. The system of claim 22, wherein the meta-document representation comprises extensible
markup language (XML) format.

24. The system of claim 22, further comprising:

analyzing means for analyzing the meta-document representation.

25. The system of claim 22, further comprising:

indexing means for indexing the meta-document representation.

26. The system of claim 21, further comprising:

progress statistics generating means for generating progress statistics regarding the
processing of the at least a portion of the information content.

27. The system of claim 26, further comprising:

progress statistics transmitting means for transmitting the progress statistics to a third
work queue.

28. The system of claim 21, wherein the first work queue and the second work queue share access to a central data structure.

29. The system of claim 28, wherein access is shared via a CORBA service.

30. The system of claim 28, wherein the data structure represents at least one of a metrics history and taxonomy regarding the information content.

31. A processor readable medium comprising processor readable code embodied therein for causing a processor to process data in a distributed architecture, the medium comprising:

- information content gathering code that causes a processor to gather information content from at least one repository according to a predetermined schedule;
- registering code that causes a processor to register the information content;
- assigning code that causes a processor to assign the information content at least one document identifier;
- work request transmitting code that causes a processor to transmit at least one work request regarding at least a portion of the information content to a first work queue;
- work request processing code that causes a processor to process the at least one work request;
- information content transmitting code that causes a processor to transmit the at least a portion of the information content to a second work queue; and
- information content processing code that causes a processor to process the at least a portion of the information content.

32. The medium of claim 31, further comprising:

- converting code that causes a processor to convert the at least a portion of the information content to a meta-document representation of the information content.

33. The medium of claim 32, wherein the meta-document representation comprises extensible markup language (XML) format.

34. The medium of claim 32, further comprising:
analyzing code that causes a processor to analyze the meta-document representation.

35. The medium of claim 32, further comprising:
indexing code that causes a processor to index the meta-document representation.

36. The medium of claim 31, further comprising:
generating code that causes a processor to generate progress statistics regarding the processing of the at least a portion of the information content.

37. The medium of claim 36, further comprising:
progress statistics transmitting code that causes a processor to transmit the progress statistics to a third work queue.

38. The medium of claim 31, wherein the first work queue and the second work queue share access to a central data structure.

39. The medium of claim 38, wherein access is shared via a CORBA service.

40. The medium of claim 38, wherein the data structure represents at least one of a metrics history and taxonomy regarding the information content.